R E P O R T R E S U M E S

ED 015 658 24 EM 004 057

AN EVALUATION OF THE EFFECT OF AN IMMEDIATE FEEDBACK DEVICE USED WITH TYPICAL COLLEGE CLASSROOM TESTS. FINAL REPORT.

BY- GRAY, ROBERT T.
SAN DIEGO STATE COLL., CALIF.

REPORT NUMBER BR-6-8156 PUB DATE JUL 67
CONTRACT OEC-4-7-068156-0232
EDRS PRICE MF-$0.25 HC-$1.56 37P.

DESCRIPTORS- *FEEDBACK, *CLOSED CIRCUIT TELEVISION, *ACHIEVEMENT TESTS, *ATTITUDES, *EXPERIMENTS, INSTRUCTIONAL TELEVISION, ACADEMIC ABILITY, SEX (CHARACTERISTICS), REINFORCERS, ANSWER GUARD FEEDBACK UNIT

PURPOSE OF THIS STUDY WAS TO EVALUATE THE EFFECTS OF AN IMMEDIATE FEEDBACK DEVICE ON LEARNING, RETENTION, AND ATTITUDES OF 264 COLLEGE STUDENTS IN A BEGINNING PSYCHOLOGY COURSE. LECTURES WERE PRESENTED ON CLOSED CIRCUIT TELEVISION. A CONTROL GROUP USED THE DEVICE ONLY TO PREVENT COPYING OF THEIR ANSWERS BY OTHER STUDENTS. NO SIGNIFICANT DIFFERENCES WERE FOUND IN POST TEST ACHIEVEMENT SCORES, NOR DID SEX OR ABILITY DIFFERENCES HAVE ANY EFFECT. STUDENT ATTITUDES WERE FAVORABLE. VALUE OF THE DEVICE SEEMED TO LIE IN SAVING TEST REVIEW TIME. (LH)

ED015658

BR 6-8156

A

PA 24

EM004057

FINAL REPORT
Project No. 6-8156
Contract No. OEC-4-7-068156-0232

AN EVALUATION OF THE EFFECT OF AN IMMEDIATE FEEDBACK DEVICE USED WITH TYPICAL COLLEGE CLASSROOM TESTS

July 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

An Evaluation of the Effect of An Immediate Feedback Device Used With Typical Classroom Tests

Project No. 6-8156
Contract No. OEC-4-7-068156-0232

Robert T. Gray

July 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

San Diego State College Foundation
San Diego State College
San Diego, California

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EM 004 057

I. Introduction

The value to learning of immediate knowledge of results is well-known. The present study undertook to evaluate the effect on learning and retention of a simple device for providing students with immediate knowledge of the correctness of their responses to the items of their regular classroom tests. If effective, this inexpensive device, called the Answer Guard feedback unit, could provide an important instructional resource for teachers at all levels as it provides a highly individualized learning situation for each pupil at no expenditure of the teacher's classroom time.

The Problem

The problem of the study was to evaluate the effect of the use of the Answer Guard feedback unit with regular college classroom tests on (1) achievement as measured by the final test and on (2) student attitude toward the classroom testing program. It was proposed also to obtain student impressions of the value of the feedback device to their learning of the course content.

Related Literature

The section on learning in the Encyclopedia of Educational Research (3:758) states that knowledge of error and success is one of the most effective reinforcers known for the college student learning under instruction. One of the most firmly established facts about reinforcement of learning is that the effect of any reinforcing event upon a response depends critically upon the delay between them (4, 5, 6, 7, 8, 9, 10, 11). According to a study by Goss (2), even a few seconds' delay between occurence of a response and reinforcement may mean the difference between maximal learning and no learning whatever.

Skinner (8) and others have pointed out that, in spite of this knowledge, typical classroom procedures are often almost maximally ineffective with respect to the delay of reinforcement. This seems to be especially true of testing activities, where the reinforcement is always delayed until the test is over and often until the papers have been scored by the teacher and returned a day or so later.

These facts clearly indicate that any teaching procedure or device which reduces this delay in reinforcement offers great promise of increased learning by students. The present study proposed to thoroughly evaluate the effectiveness of a simple, inexpensive device for this purpose.

Since the teaching technique proposed for evaluation in the study had not previously been possible, due to the newness of the device upon which it is based, no research literature directly related to the proposed study was available.

The Hypotheses

1. College students who are provided with immediate knowledge of the correctness of their responses to items of regular classroom tests achieve final examination scores not significantly different from those of students provided with posttest review and discussion.

2. College students who are provided with immediate knowledge of the correctness of their responses to items of regular classroom tests rate the quality of the testing program at a level not significantly different from that of students provided with posttest review and discussion.

The first hypothesis was investigated for all groupings of students resulting from fully crossing sex, two categories of ability, and two class meeting times. The second hypothesis was investigated for the four groups defined by sex of students and class meeting time.

3. At the time of the final examination men and women college students express equivalent attitudes toward the use of an immediate feedback device with their regular classroom tests.

Sex differences in attitudes toward the use of the feedback device were studied by means of the above hypothesis. Overall student attitudes toward the use of the device were studied on a descriptive basis without formal hypotheses as only the experimental classes used them.

II. The Method

At San Diego State College a general education course in psychology is required of all students. Most students take this course (Psychology I) during their freshman year. Approximately twenty-five sections are offered each semester. Twelve of these sections of Psychology I utilize closed-circuit television to present all lectures. These classes are ideally suited to experimental research in education because they (1) contain a representative sample of freshman students at San Diego State College, (2) receive a common program of lectures via closed-circuit television, and (3) take common examinations. The present study utilized the twelve Psychology I classes which were offered by television during the fall semester of the 1966-67 academic year.

The Sample

The television sections of Psychology I are regularly scheduled so that six meet at nine o'clock in the morning on Monday, Wednesday, and Friday and the other six at eleven on the same days. The classes are limited to no more than fifty members and usually enroll at least thirty-five. Six graduate students pursuing a masters degree in the Psychology Department are assigned as teaching assistants to the television sections. Each of them is responsible for one nine and one eleven o'clock section. The sections are numbered from 1 to 12 with one teaching assistant assigned to Section 1 at nine and Section 7 at eleven, the next teaching assistant to Section 2 at nine and Section 8 at eleven and so on to the sixth teaching assistant who was assigned to Section 6 at nine and Section 12 at eleven.

Sections 1, 3, and 5 were designated as the nine o'clock experimental group and 2, 4, and 6 as the nine o'clock control group. For the six sections which met at eleven o'clock Sections 7, 9, and 11 were designated as experimental and Sections 8, 10, and 12 as control. In this way each teaching assistant was assigned to one experimental and one control section.

Because no instructor's name was given in advance for any of the television Psychology I classes, it was assumed that there was no systematic tendency for students to enroll in any particular section offered at a given hour. To eliminate any bias associated with enrollment at the nine or eleven o'clock periods and because of noncomparable examinations, comparisons were made only between the three experimental and three control sections offered during the same period.

Although all students in the experimental sections were subjected to the experimental treatment, the actual experimental comparisons were only carried out for eleven male and eleven female students from each class section, drawn at random from those who completed all of the examinations given during the course of the experiment. The final sample, then, consisted of sixty-six experimental and sixty-six control subjects for the nine o'clock classes and a like number for the eleven o'clock period, a total of two hundred sixty-four subjects for the complete study.

The Experimental Design

Because of the large number of students and sections of Psychology I available for the study and because of their essentially random assignment to classes within each of the two time periods for which television sections were scheduled, it was possible to utilize the posttest-only experimental design. This design, which is one of the true experimental designs listed by Campbell and Stanley (1) in the Handbook of Research on Teaching basically involves the assignment of subjects by random means to two groups, the application of an experimental treatment to one of the two groups, followed by measurement for both groups of the dependent variable. Any significant observed difference between the two groups on this posttesting can be taken as evidence of the differential effect of the experimental treatment.

To reduce the likelihood that students in the nine o'clock classes would transmit test answers to the eleven o'clock students, separate forms of each test were given during the two time periods. Because of this difference in tests, a separate experimental design was set up for the nine and eleven o'clock classes. Thus, three of the sections meeting at nine o'clock constituted an experimental group and were compared to the other three classes meeting at the same time. This pattern was repeated for the six classes meeting at eleven o'clock with a different set of tests being used both in presenting the experimental treatment and in making the posttest comparisons.

The experimental designs were balanced by choosing eleven male and eleven female students at random from each of the twelve classes involved in the experiment for the statistical analysis.

The Experimental Treatment

All of the lectures used in the experimental and control classes of Psychology I were presented over closed-circuit television to sets located in the individual classrooms. The

original video-tapes of the lectures were made by Dr. Lark O. Daniel, Professor of Psychology at San Diego State, who is charged with overall responsibility for the television sections of Psychology I. Each of the lectures was followed by a class discussion of the subjects covered, with the teaching assistant leading the discussion and answering questions. A schedule of the lectures and examinations given during the experiment is included as Appendix A of this report.

The experimental treatment consisted of the provision of immediate knowledge of the correctness of responses made by members of the experimental class to items of the four classroom tests given during the semester and prior to the final examination. This immediate feedback was provided by means of a testing accessory known as the Answer Guard.

The Answer Guard is a clear plastic overlay, one-twelfth of an inch thick, which fits over the standard IBM test answer sheet. It is slotted over each of the 750 answer positions provided by the answer sheet so that answer marks may be made through it. Its original purpose was to provide security for the responses of individual students in large group testing situations. The refraction of light passing through the sides of the answer slots makes it impossible to view answer marks from any position other than directly above the answer sheet.

The feedback device used in the study consisted of a unit made up of a standard IBM answer sheet placed between two Answer Guards and then placed over another answer sheet on which the correct answer spaces had been overprinted with red marks. A cardboard backing sheet to support the bottom answer sheet completed the unit, which was held together by two rubber bands stretched between diagonally opposite corners. A plan view and cross-sectional view of the feedback unit are shown as Figure I and Figure II in Appendix F.

Students in the experimental classes responded to items of their Psychology I tests by punching through the uppermost answer sheet with a stylus. This perforation of the upper answer sheet revealed the corresponding answer space of the bottom answer sheet. If the choice were correct the bottom answer sheet would reveal its overprinted red mark. If the red mark was not revealed, the student knew immediately that the belief which had led him to select the answer choice was not correct and should be reconsidered. Upon the student's selection of another answer to the item the process was repeated until the correct choice was finally made, and verified by the appearance of the red mark.

Because of the reluctance of the professor in charge of the Psychology I classes to allow partial credit for items answered correctly on the second or third attempt it was necessary to count items which revealed the red mark in one punch as right and those requiring more than one punch as wrong, regardless of the number of such additional punches. As a result it appeared that many students failed to thoughtfully consider their answer choices after the first one and, rather, just punched at random to reveal the correct answer. If this was actually the case, it might have reduced to some degree the value of the immediate feedback in increasing learning and retention.

The four classroom tests with which the experimental treatment was provided were given at the conclusion of each of the four units which make up the Psychology I course content. Each of the tests was made up of forty items. Thirty of the items were common to all of the sections meeting at the same hour. Thus all of the nine o'clock sections used one set of thirty items in common. An additional ten items were made up individually for each section by the teaching assistant in charge. Students in the experimental sections all responded on the immediate feedback device while students in the control classes marked their answer sheets through a single Answer Guard, provided solely for the purpose of eliminating any chance for students to copy each others' answers.

Some students in the experimental classes expressed concern at not being able to change answers once punched. They quickly adjusted to the new system, however, and only a few maintained this concern to the end of the semester. Responses to a questionnaire which sought an evaluation of the use of the feedback technique are given in a later section of this report and give a clear indication of the extent of this tendency.

Boxes containing preassembled Answer Guard units and all necessary materials were delivered to teaching assistants just prior to each testing period and the testing was carried out by them in the classes for which they were responsible.

The Data Collection

The regular Psychology I final examination served as the major criterion measure for evaluating the experimental effect of immediate answer feedback during the administration of four previous unit tests. This examination consisted of one hundred four-choice multiple response items. A separate form was administered to the nine o'clock and eleven o'clock class sections.

Items of the criterion measure were based on the same subject matter as the items of the unit tests with which the independent variable --- immediate feedback --- was introduced. Both experimental and control classes were tested under identical conditions and used Answer Guards to insure complete security from answer copying. All students were provided with printed instructions for using the Answer Guard unit. Classes using the feedback unit received different instructions from those using the single Answer Guard for security purposes only. Copies of these instructions are shown in Appendix B of this report.

Student reactions to the provision of immediate feedback and to the use of the Answer Guard as a security device were obtained by means of a brief questionnaire administered with the final examination. Items relating to immediate feedback were omitted from the form used with the control classes as they were not exposed to it. Copies of the questionnaire forms used are included in Appendix C.

Methods of Analysis

The first set of hypotheses -- those dealing with experimental differences in the performance of various student groups on the Psychology I final examination -- were tested by means of analysis of variance comparisons. Comparisons were carried out within all six groups resulting from a complete blocking on sex, class meeting time, and two categories of ability. The two ability categories were simple divisions of the students in each sex-class breakdown into upper and lower halves with respect to performance on the first unit test. Thus, of the eleven boys from each class, six were classified as high ability and five as low on the basis of their relative standings on the test.

The responses of experimental and control groups to identical questions of the questionnaire concerning their evaluation of the Psychology I testing program were compared by means of appropriate chi-square techniques. Questionnaire items seeking reactions to the different testing procedures used in the experimental and control classes were organized for presentation on a strictly descriptive basis except for comparisons of subgroups within treatment groups, for which appropriate non-parametric techniques were employed.

III. Results

Results of the experimental comparisons provided for in the design of the study are detailed in Tables 1 through 8 in Appendix D. Tables containing purely descriptive data are in Appendix E.

A preliminary analysis of variance comparison of the dependent variable means for the three separate Psychology I classes falling within each treatment group revealed no significant differences. The twelve classes involved in the study were therefore pooled into four groups, by treatment (immediate feedback versus review-discussion) and by hour (nine and eleven o'clock). By virtue of the original random selection of eleven male and eleven female students from each of the twelve classes, the final four comparison groups each contained 66 students. It was assumed, because the classes were poolable taken as a whole, that their subgroups by sex and ability would also be poolable.

The primary hypothesis of the study --- that provision of immediate knowledge of the correctness of responses to items of regular college classroom tests would not result in final examination achievement scores significantly different from those resulting from posttest review and discussion --- was accepted for all comparison groups. Analysis of variance comparisons of mean final scores of experimental and control students were made for all groups resulting from crossing sex, ability, and time of day for eleven male and eleven female students picked at random from each of the twelve classes involved in the study. None of the resulting F-ratios was significant at the .05 level of confidence.

Because of the anonymous nature of the student reaction questionnaires it was necessary to use all of them, rather than just those from the random sample used in the other statistical analyses. Comparison of the experimental and control groups with respect to their evaluation of the Psychology I testing program revealed only one significant difference --- female students in the nine o'clock control classes had more superior and above average ratings than females in the experimental classes. The second hypotheses presented on page 2 was therefore accepted on an overall basis but rejected for the female members of the nine o'clock sections.

Although the provision of immediate feedback did not produce significantly higher final examination scores than posttest review and discussion, seventy-one per cent of the experimental students stated that they felt their learning and

retention of the Psychology I course material had been improved. Comparison of the responses of males and females within class periods revealed virtually no differences, but an unexpectedly significant difference was found between the nine and eleven o'clock groups. Sixty-two per cent of the nine o'clock group claimed improvement, with thirty-three per cent indicating great improvement, against seventy-nine per cent of the eleven o'clock group, with forty-five per cent greatly improved. Table 9 in Appendix E contains details of this analysis.

Both experimental (81 per cent) and control (88 per cent) classes recommended strongly that the use of Answer Guards to conceal student responses from view be continued. Experimental comparisons for all groups by sex and class period revealed only one significant difference --- a greater proportion of female students in the eleven o'clock control group recommended continued use. Complete data are given in Table 8 in Appendix D.

When asked what system of feedback scoring they preferred, students in the experimental groups gave only minor support to the system used in the experiment. Only thirty per cent preferred that only answers punched correctly in one attempt be given credit and all others no credit. Although this scoring procedure was necessary for the study, because of the wishes of the instructional staff, it appeared from examination of test papers that it led to random punching after an initial wrong choice with a resultant loss of effectiveness of the feedback device. A majority (56 per cent) of the students expressed preference for a scoring sustem which would give answers punched correctly the first time full credit, ones punched correctly in two tries half credit, and ones punched correctly in three tries one-quarter credit. A smaller group (14 per cent) indicated a preference for some other system of scoring.

Although too lengthy for inclusion in this report the written comments on the student reaction questionnaires gave valuable insight into student impressions of the use of the Answer Guard unit for feedback and to prevent copying. Those comments concerning the feedback unit were predominantly favorable, but a significant number expressed frustration at knowing that they had missed items, especially if a number were missed in sequence. A number also commented on the need for partial credit for second and third attempts in order to encourage thoughtful reconsideration of alternatives.

Concerning the function of the Answer Guard to prevent copying, only a few comments were negative. Most of these expressed a belief that cheating on tests is not a problem to be concerned about.

Tables detailing all the objective results of the study are included in the various appendixes at the end of this report.

IV. Discussion

The effectiveness of the feedback device may have been seriously impaired by the failure of the scoring system to provide a credit incentive for making second and third choices of responses to items as thoughtfully as the first. Additionally, the relative brevity of the unit tests with which the feedback was provided --- thirty items --- may have reduced the advantage of the feedback system over posttest review and discussion.

The internal validity of the experimental design appears to have been exceptionally high. Few opportunities for college classroom research could offer the random assignment of students to classes, identical television lectures, and common testing available for this study.

V. Conclusions and Implications

Basically, it may be concluded from the study that provision of immediate knowledge of the correctness of responses to items of regular classroom tests at the college level does not differ significantly in its effect on final examination scores from that of posttest review and discussion. However, because the provision of immediate feedback through the use of the Answer Guard unit requires no additional classroom time, being limited entirely to the testing period, it does represent a contribution to learning.

The study failed to reveal any significant sex or ability differences in the effect of the feedback device. This was contrary to expectations that the female students would be less favorably disposed toward a mechanical device.

One implication of the study seems inescapable. The advantage of the use of the Answer Guard feedback unit over the more conventional technique of posttest review and discussion lies only in the conservation of classroom time, not in increased performance on the final examination.

VI. Summary

The study was an evaluation of the effect on learning and retention of a device for providing college students with immediate knowledge of the correctness of their responses to the items of their regular classroom tests. The feedback unit consisted of two IBM test answer sheets separated by two clear plastic shields known as Answer Guards. The lower answer sheet has the correct answer spaces overprinted with red spots which are revealed when the upper answer sheet is perforated. Thus, the examinee is provided with immediate knowledge of the correctness of each answer punch he makes. The Answer Guard also makes answer marks invisible except from directly above the answer sheet, thus preventing the copying of answers.

The sample for the study consisted of students enrolled in twelve sections of Psychology I at San Diego State College during the fall semester of 1962. All sections received their classroom lectures by means of closed circuit television, took common examinations, and contained students enrolled at random from a pool of students taking the course.

Students in the six experimental classes used the feedback device in taking their four unit examinations during the semester and were thereby furnished with immediate feedback concerning the correctness of each answer selection. Students in the six control classes took their unit examinations in the normal manner except that Answer Guards were placed over their answer sheets to prevent any copying and they received posttest review and discussion.

Eleven male and eleven female students from each of the twelve classes involved in the study were drawn at random for the analysis of results. When the final examination scores of the hundred and thirty-two experimental students were compared with the same number of control students no significant differences were found. Further analysis failed to reveal any significant sex or ability differences in the effect of the feedback device. Student reaction to the use of the feedback device was generally favorable with a few students expressing frustration at finding that answers confidently expected to be correct were wrong, especially sequences of such answers. Almost all students favored the use of the Answer Guard unit to eliminate copying during tests.

It was concluded that the advantage of providing immediate feedback of results during classroom tests at the college level lies not in the improvement of final examination scores but, rather, in saving the classroom time normally devoted to the review and discussion of the tests.

VII. References

1. Gage, N. L. Handbook of Research On Teaching. Chicago: Rand McNally & Company. 1963. 1218p.

2. Goss, Albert E. "Stimulus-response Analysis of the Interaction of Cue-producing and Instrumental Responses," Psychological Review. 62, 1955. p. 20-31.

3. Harris, Chester W. Encyclopedia of Educational Research. New York: The Macmillan Company. 1960. 1564p.

4. Hull, Clark L. Principles of Behavior. New York: Appleton. 1943. 442p.

5. Kimble, Gregory A. "Conditioning as a Function of the Time Between Conditioned and Unconditioned Stimuli," Journal of Experimental Psychology. 37, 1947 p. 1-15.

6. Lorge, Irving, and Thorndike, Edward L. "The Influence of Delay in After-Effect of a Connection," Journal of Experimental Psychology. 18, 1935. p. 187-194.

7. Reynolds, Bradley. "The Acquisition of a Trace Conditioned Response as a Function of the Magnitude of the Stimulus Trace," Journal of Experimental Psychology. 35, 1945. p. 15-30.

8. Skinner, B. F. "The Science of Learning and the Art of Teaching," Harvard Educational Review. 24, 1954. p. 86-97.

9. Spence, Kenneth W. "The Role of Secondary Reinforcement in Delayed Reward Learning," Psychological Review. 54, 1947. p. 1-8.

10. Spence, Kenneth W. Behavior Theory and Conditioning. New Haven: Yale University Press. 1956. 262p.

11. Spooner, Alice; and Kellogg, Winthrop N. "The Backward Conditioning Curve," American Journal of Psychology. 50, 1947. p. 321-334.

Appendix A

Course Calendar for Psychology I

Course Calendar Psychology I Television Sections

Television Lecture Topics

First day of classes -- Introduction to the Course

ITV(1): Overall Objectives of the General Psychology Course

ITV(2): Nonsense in "Common Sense" Psychology

ITV(3): Explaining Behavior

ITV(4): Behind Human Behavior, Part 1: Heredity and Maturation

ITV(5): Behind Human Development, Part 2: Psychosocial Development

ITV(6): To Live Is To Learn To Live

ITV(7): Classical and Instrumental (Operant) Conditioning

ITV(8): Complex Learning

ITV(9): Teaching Machines and Programmed Learning

ITV(10): Why Do You Forget?

Test #1

ITV(11): The Wellsprings of Human Behavior

ITV(12): The Genesis of "Motor-Vation"

ITV(13): Motivation: Subconscious and Acquired

ITV(40): Frustration and Conflict

ITV(14): Emotions

ITV(15): Emotions and Personality

ITV(42): The Magnificent Mutation

ITV(20): That's the Way We're Wired

ITV(21): The Sensible Use of Our Senses

ITV(22): Seeing Is Believing

ITV(23): Factors in Perception

Test #2

ITV(24): Thinking

ITV(25): Toward Accurate Thinking

ITV(27): Intelligence: By Guess and By Golly

ITV(41): Facts, Figures, and Fiction:

ITV(26): Personality: Development and Functioning

ITV(28): Personality: The Self Concept

ITV(29): Each Man Is An Island

Test #3

ITV(16): Attorney For the Defense: You!

ITV(17): Shades of Gray

ITV(18): Psychological Therapy

ITV(30): No Man Is An Island

ITV(32): Facts, Figures, and Fiction: Sampling and Probability

ITV(33): The Dynamics of Communication

ITV(34): Too Much, Too Fast, Too Soon

ITV(35): Privileges or Immunities of Citizens

ITV(31): Automation: Problem and Promise

ITV(37): Rapprochment and Quo Vadis

Test #4

Last day of classes before final examinations

Final Examination

Appendix B

Instructions For Use of Answer Guard Units

Directions for students using the Answer Guard feedback device.

As an aid to learning, this answer sheet unit has been prepared to provide you with immediate knowledge of the correctness of each response you make to the items of this test. Each time you select an answer to a test item, mark it by punching through the top answer sheet with the styllus provided. If you have selected the correct answer a red spot will appear in the opening which has been punched. If your selection is not correct, reconsider the item and punch a new answer space. Continue this procedure until you have punched out the correct answer space and the red spot has been exposed.

The answer spaces from 46 through 60 have been prepared to give you practice in punching and recognizing the red spots. You will find that a vertical, punching stroke is more effective than a horizontal, tearing stroke.

Only items requiring a single punch to reveal the correct answer will be counted as correct. However, if you fail to reveal the correct answer in one punch, you should thoughtfully select each additional answer as this will increase your learning. A correct answer must be exposed for every item!

Directions for students using the Answer Guard.

For this test you will be marking your answers through a transparent plastic shield called an Answer Guard, which will protect them from the view of other students seated near you. Make your answer marks through the shield in the usual way, using the special pencil and making sure that they are shiny and black.

If you make an answer mark and then wish to change it, proceed as follows. First, write the number of the answer mark you wish to have erased in the opening at the upper left-hand corner of the Answer Guard. For example, if you had marked the second answer space for item 17, you would write 17-2 in the space for "answers to be erased." Next, proceed to mark the new answer. When the papers are scored, all answer marks whose numbers appear in the upper left-hand corner of the answer sheet will be erased. If you decide later that you don't want an answer mark erased after all, simply cross out its number.

Appendix C

Student Reaction Questionnaire Forms

Questionnaire for experimental sections.

As a basis for continued improvement of the Psychology I testing program, your reactions to the following questions concerning the use of Answer Guards would be of value. You need not give your name.

1. Section___ 2. Sex___ 3. Age___ 4. Estimated Psych I grade___

5. Compared to other courses I am taking this semester, the testing program in Psychology I has been

___ (a) superior.
___ (b) above average.
___ (c) average.
___ (d) below average.
___ (e) inferior.

Comments _____

6. As a result of my experience in Psychology I, I would recommend that the use of Answer Guards to conceal student responses from view be

___ (a) continued for future classes.
___ (b) discontinued.

Comments _____

7. I feel that the immediate knowledge of the correctness of my responses to test items provided by the Answer Guard unit has _____ my learning and retention of the material taught in Psychology I.

___ (a) greatly improved
___ (b) slightly improved
___ (c) had no effect on
___ (d) slightly reduced
___ (e) greatly reduced

Comments _____

8. With regard to scoring tests taken with the immediate feedback feature of the Answer Guard unit I would prefer that

___ (a) only answers punched correctly the first time be given any credit.
___ (b) answers punched correctly the first time be given full credit, ones punched correctly in two tries be given half credit, and ones punched correctly in three tries be given one-quarter credit.
___ (c) another procedure for scoring be used.

Questionnaire for control sections.

As a basis for continued improvement of the Psychology I testing program, your reactions to the following questions concerning the use of Answer Guards would be of value. You need not give your name.

1. Section___ 2. Sex___ 3. Age___ 4. Estimated Psych I grade___

5. Compared to other courses I am taking this semester, the testing program in Psychology I has been

___ (a) superior.
___ (b) above average.
___ (c) average.
___ (d) below average.
___ (e) inferior.

Comments _____

_____

_____

6. As a result of my experience in Psychology I, I would recommend that the use of Answer Guards to conceal student responses from view be

___ (a) continued for future classes.
___ (b) discontinued.

Comments _____

_____

Appendix D

Tables for Experimental Results

Table 1

TESTS OF POOLABILITY FOR CLASSES WITHIN TREATMENT GROUPS

Groups	N	$\bar{X}$	SD	Variance Between	Variance Within	F	Hyp.	P
Experimental (9:00)								
Class 01	22	71.8	9.3					
Class 03	22	76.6	8.4	207.5	78.3	2.65	Acc.	>.05
Class 05	22	71.0	8.8					
Control (9:00)								
Class 02	22	72.2	10.3					
Class 04	22	71.5	8.3	3.0	89.6	.03	Acc.	>.05
Class 06	22	72.2	9.7					
Experimental (11:00)								
Class 08	22	77.9	10.6					
Class 10	22	78.0	8.1	31.5	81.8	.38	Acc.	>.05
Class 12	22	75.9	8.2					
Control (11:00)								
Class 07	22	77.8	9.3					
Class 09	22	78.5	6.9	99.9	74.7	1.34	Acc.	>.05
Class 11	22	74.5	9.5					

Table 2

COMPARISON OF FINAL EXAMINATION SCORES OF EXPERIMENTAL (FEEDBACK) AND CONTROL (NON-FEEDBACK) GROUPS DIVIDED BY SEX AND ABILITY

Groups	N	$\bar{X}$	SD	Variance Between	Variance Within	F	Hyp.	P
High Male (9:00)								
Experimental	16	75.25	7.82	.78	65.30	.01	Acc. > .05	
Control	16	75.56	8.33					
Low Male (9:00)								
Experimental	17	71.59	8.97	169.88	86.43	1.97	Acc. > .05	
Control	17	67.12	9.61					
High Female (9:00)								
Experimental	16	77.62	9.60	19.53	74.49	.26	Acc. > .05	
Control	16	76.06	7.54					
Low Female (9:00)								
Experimental	17	68.41	7.70	11.76	70.51	.16	Acc. > .05	
Control	17	69.59	9.04					

Table 3

COMPARISON OF FINAL EXAMINATION SCORES OF EXPERIMENTAL (FEEDBACK) AND CONTROL (NON-FEEDBACK) GROUPS DIVIDED BY SEX AND ABILITY

Groups	N	$\bar{X}$	SD	Variance Between	Variance Within	F	Hyp. P
High Male (11:00)							
Experimental	16	78.44	11.62	.28	106.30	.00	Acc. > .05
Control	16	78.25	8.81				
Low Male (11:00)							
Experimental	17	72.71	6.85	10.62	67.24	.16	Acc. > .05
Control	17	71.59	9.35				
High Female (11:00)							
Experimental	16	81.38	8.43	13.78	58.57	.24	Acc. > .05
Control	16.	82.69	6.79				
Low Female (11:00)							
Experimental	17	76.76	6.73	11.77	40.60	.29	Acc. > .05
Control	17	75.59	5.99				

Table 4

COMPARISON OF EXPERIMENTAL (FEEDBACK) AND CONTROL (NON-FEEDBACK) GROUPS WITHIN SEXES

Groups	N	$\bar{X}$	SD	Variance Between	Variance Within	F	Hyp.	P
Male (9:00)								
Experimental	33	75.48	9.76	7.33	93.39	.088	Acc. >	.05
Control	33	74.82	9.57					
Female (9:00)								
Experimental	33	79.00	7.84	.02	56.98	.000	Acc. >	.05
Control	33	79.03	7.25					
Male (11:00)								
Experimental	33	73.36	8.51	76.38	84.74	.910	Acc. >	.05
Control	33	71.21	9.85					
Female (11:00)								
Experimental	33	72.88	9.73	.37	86.50	.010	Acc. >	.05
Control	33	72.73	8.85					

Table 5

COMPARISON OF FINAL EXAMINATION SCORES OF EXPERIMENTAL (FEEDBACK) AND CONTROL (NON-FEEDBACK) GROUPS WITHIN ABILITY GROUPS

Groups	N	$\bar{X}$	SD	Variance Between	Variance Within	F	Hyp.	P
High (9:00)								
Experimental	32	76.44	8.70	6.24	68.40	.09	Acc. >	.05
Control	32	75.81	7.82					
Low (9:00)								
Experimental	34	70.00	8.39	46.12	78.18	.59	Acc. >	.05
Control	34	68.35	9.27					
High (11:00)								
Experimental	32	79.91	10.09	5.06	83.43	.06	Acc. >	.05
Control	32	80.47	8.06					
Low (11:00)								
Experimental	34	74.74	7.00	22.37	56.47	.40	Acc. >	.05
Control	34	73.59	8.00					

Table 6

COMPARISON OF FINAL EXAMINATION SCORES OF TOTAL EXPERIMENTAL (FEEDBACK) AND CONTROL (NON-FEEDBACK) GROUPS

Groups	N	$\bar{X}$	SD	Variance Between	Variance Within	F	Hyp.	P
9:00								
Experimental	66	73.12	9.07	43.75	84.62	.52	Acc.	> .05
Control	66	71.97	9.32					
11:00								
Experimental	66	77.24	8.96	3.37	77.85	.04	Acc.	> .05
Control	66	76.92	8.69					

Table 7

COMPARISON OF STUDENT EVALUATIONS OF THE TESTING PROGRAM

Groups (9:00)	Question: Compared to other courses I am taking this semester, the testing program in Psychology I has been superior.	above average.	average.	below average.	inferior.	χ^2*
Male						
Exper.	9	24	7	11	7	
Control	3	23	26	8	3	.30
Female						
Exper.	3	23	20	16	2	
Control	9	26	31	6	1	6.12 Sig. P <.05
Total						
Exper.	12	47	27	27	9	
Control	12	49	57	14	4	4.63 Sig. P <.05
Groups (11:00)						
Male						
Exper.	10	25	13	7	1	
Control	1	18	19	5	4	1.71
Female						
Exper.	12	28	18	7	3	
Control	5	28	28	5	2	.09
Total						
Exper.	22	53	31	14	4	
Control	6	46	47	10	6	.41

*based on number above and below average.

Table 8

STUDENT RECOMMENDATIONS FOR USE OF ANSWER GUARD FOR TEST SECURITY

	Question: As a result of my experience in Psychology I, I would recommend that the use of Answer Guards to conceal student responses from view be		
Groups (9:00)	continued for future classes.	discontinued.	χ^2
Male			
Exper.	44	11	.14
Control	48	10	
Female			
Exper.	51	14	2.32
Control	60	8	
Total			
Exper.	95	25	1.83
Control	108	18	
Groups (11:00)			
Male			
Exper.	50	4	2.62
Control	36	8	
Female			
Exper.	51	17	9.80 Sig. P<.01
Control	66	4	
Total			
Exper.	101	21	2.17
Control	102	12	

Appendix E

Tables for Descriptive Results

Table 9

STUDENT EVALUATION OF THE EFFECT OF FEEDBACK

	Question: I feel that the immediate knowledge of the correctness of my responses to test items provided by the Answer Guard unit has _____ my learning and retention of the material taught in Psychology I.					
Groups (9:00)	greatly improved	slightly improved	had no effect on	slightly reduced	greatly reduced	χ^{2*}
Male	15	23	15	2	2	.18
Female	9	26	21	5		
Total	24	49	36	7	2	4.77** Sig. P<.05
Groups (11:00)						
Male	17	31	8	1		.40
Female	27	24	15		2	
Total	44	55	23	1	2	

*number improved versus number reduced.

**total 9:00 versus total 11:00.

Table 10

STUDENT PREFERENCES FOR FEEDBACK SCORING

	Question: With regard to scoring tests taken with the immediate feedback feature of the Answer Guard unit I would prefer that		
Groups (9:00)	(a)	(b)	(c)
Male	27	24	7
Female	9	40	11
Total	36	64	18
Groups (11:00)			
Male	12	35	8
Female	23	35	9
Total	35	70	17

(a) only answers punched correctly the first time be given any credit.

(b) answers punched correctly the first time be given full credit, ones punched correctly in two tries be given half credit and ones punched correctly in three tries be given one-quarter credit.

(c) another procedure for scoring be used.

Appendix F

Plan and Cross-Sectional Views of the Feedback Unit

Figure I

PLAN VIEW OF THE FEEDBACK UNIT

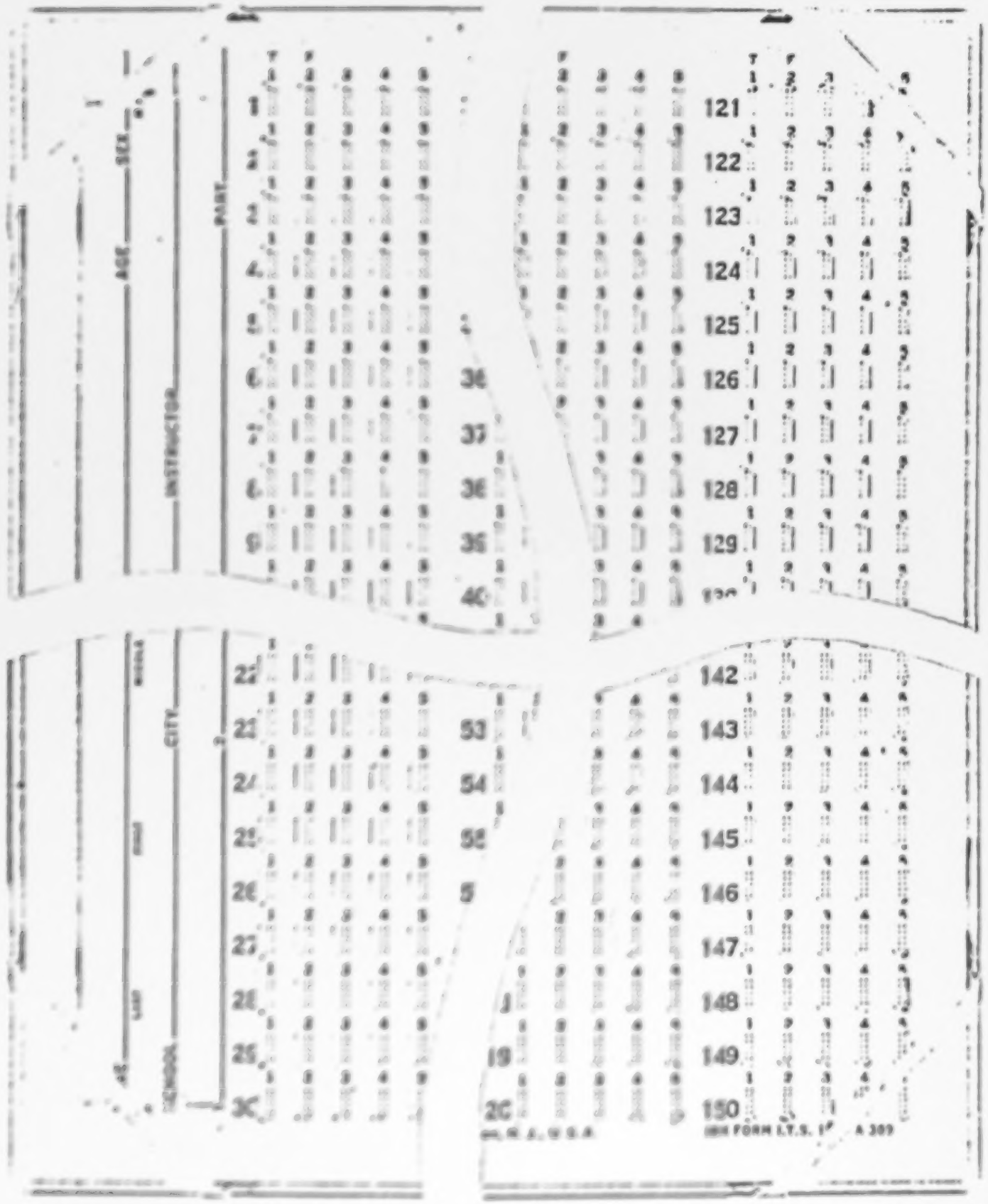

Figure II

CROSS-SECTIONAL VIEW OF THE FEEDBACK UNIT